

Interactive comment on “Advancing land surface model development with satellite-based Earth observations” by Rene Orth et al.

Anonymous Referee #3

Received and published: 28 February 2017

This manuscript by Rene Orth and fellow co-authors is a follow-on companion paper to the Orth et al. (2016) article, referred to in this manuscript as “O16”. The new components in this HESS manuscript expanded over the O16 paper include evaluating their O16 simulations with geostationary satellite-based land surface temperature (LST) observations and exploring more the number of metrics and variables in combination with the LST observations. In the original O16 paper, the authors performed a parameter sensitivity analysis with 2300 offline HTESSEL simulations involving 6 in-situ observation sites (2 in Finland, 2 in Switzerland, and 2 in southern Italy). From these simulations, the authors pulled the best performing 25 (ranked) parameter datasets and then randomly selected another (different) 25 parameter datasets, along with the default parameter dataset. They identified in O16 six HTESSEL parameter types that the model may be most sensitive to and impacted by changes to. These same 6 are

C1

used and evaluated in this study.

Major comments:

- 1) Please specify in this manuscript whether the exact 51 parameter datasets that were selected and used in the related offline HTESSEL simulations in O16 are the same as used in this study.
- 2) In section 2.3, are these performance measures applied the exact same way as in O16? The descriptions are different between the two papers. Also, is the calculated bias literally the difference between the average of the observations and average of the model simulation output, using the same overlapping years (e.g., 8 years of LST data)? Please describe further in your paper.
- 3) For the 11 soil moisture sites used in the evaluation, do they overlap at all with the original 6 sites used in the O16 calibration and evaluation paper? Or are these in addition to those 6 sites. Please specify and add in your figure 1 plot the location of those sites (or add a new figure highlighting the locations of the observations, including the 400+ streamflow gage sites).
- 4) In subsection 2.1.1, please further describe how the subgrid tiles of the HTESSEL model are organized (e.g., using top 5 dominant vegetation types per gridcell), or if just the dominant vegetation type tile is considered per gridcell.
- 5) In Figure 1, the authors consider grid cells where land cover is > 80%. However, most of the forested regions of “high vegetation” types look to be screened out from the analysis. Please explain if and how this might tie back to the LST screening process outlined in section 2.5 and how the vegetation tile process relates to it.
- 6) Figure 2 is essentially the same as Figure 3 in Orth et al. (2016) but with adding the LST reference dataset and focusing on just the 51 parameter datasets. In the description for this Figure 2, the authors however state in lines 13-15 on page 9 that the hydrology-based reference datasets are “sensitive mostly to stomatal conductivity”.

C2

Are you referring here to the minimum stomatal resistance parameter (minStoRes), which would match the results of Fig. 2, or the skin conductivity (SkinCond) parameter, which would not be right according to the results shown? Similarly, what is stated for the skin temperature plots in Fig. 2 that the “performance is especially sensitive to minimal stomatal conductivity parameter”. However, in Figure 2 the LST is highly sensitive to the “SkinCond” parameter. Please make the qualitative analysis description in these lines much more clearly reflect the results of Figure 2, and using the correct names of the parameters with the highest sensitivity.

7) In lines 35-37 on page 9, it is confusing when the authors state: “The many dark colored fields in Figure 3 indicate that different parameter perturbations perform best against different datasets.”. However just above that statement, the authors say that the light colors indicate parameters that perform well against the reference datasets. Please rephrase these sentences to reduce any conflicting statements of what the light colors refer to.

8) The authors refer on line 2 of page 11, “i.e., the more white color there is in Figure 1)” . . . which areas of white color are you referring to? There are no white areas to refer to. Here you are referring to Figure 1 and on lines 2 and 4 of page 11. Please rewrite this paragraph to make it clearer and make sure you are referring to the correct figures(s).

9) One aspect of the geostationary satellite-based LST observations used as a reference dataset in this study that is hardly mentioned or discussed is what are the actual errors and uncertainty of this observed LST dataset itself. How does this factor in to the model evaluation (e.g., bias metric)? Please address this further in the paper. For example, if the observed LST dataset is heavily biased, if you were to calibrate the model to such biased observations, you could constrain the model to this bias.

10) In terms of the model spin-up period, 1-year spinup is not usually sufficient, especially for soil moisture and temperature in desert regions, where it has been shown

C3

that it can take more than 20 years for such locations to reach equilibrium. Please consider running a longer spinup period and rerun the analysis for the LST evaluation. Would the longer spinup affect the model LST bias in the bare soil regions, like Sahara desert.

Minor Comments:

1) In subsection 3.2.2., please provide additional details on the “E-OBS” dataset, which is supposedly used in validating the coupled forecasts. 2) On page 4, line 5, need to place a period after “. . .Balsamo et al., 2011)” reference.

3) In Figure 1, please define in the figure caption and in the main text what “DYN” stands for (which is only found on the “y-axis” labels for the bottom two rows of plots).

4) Also regarding the Figure 1 description in Section 4, please confirm in the text the “direction of bias”, e.g., obs – model (default parameters, based on the bias definition in section 2.3). This is to help clarify when explaining in the two top rows of Figure 1 that the model is greatly underestimating the diurnal range in the more vegetated regions and overestimating in desert regions, like the Sahara.

5) What does the gray areas indicate in Figure 1?

6) On page 10, lines 32-36, this appears to be a run-on sentence. Please consider making this sentence into two with the start of the new sentence begin on line 30, “The best-performing parameter sets . . .”.

7) Also for the point #1 starting on line 30, it is difficult to follow. Please try to more concisely explain your results and point made here relative to Figure 4. Which “two validation metrics” are you referring to?

8) On page 11, line 33, make “were” to “where”, as “where both the mean LST . . .”.

9) For the LST data background, is what is described in section 2.5 the same dataset as described in section 3.2.1 on page 8? If so, please consider combining into one of

C4

the given sections.

10) For the Section 3.1, please add more to the section on the WFDEI forcing dataset.

References: Orth, R., E. Dutra, and F. Pappenberger (2016), Improving weather predictability by including land-surface model parameter uncertainty. *Mon. Weather Rev.*, 144(4), 1551-1569

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-628, 2016.